DOCUMENT RESUME

ED 466 227 IR 021 298

AUTHOR	Vrabic, Gorazd; Simon, Bernd
TITLE	Learning Resource Catalogue Design of the UNIVERSAL Brokerage Platform.
PUB DATE	2001-00-00
NOTE	7p.; In: ED-Media 2001 World Conference on Educational Multimedia, Hypermedia & Telecommunications. Proceedings (13th, Tampere, Finland, June 25-30, 2001); see IR 021 194.
AVAILABLE FROM	Association for the Advancement of Computing in Education (AACE), P.O. Box 2966, Charlottesville, VA 22902 ($40, AACE members; $50, nonmembers). Tel: 804-973-3987; Fax: 804-978-7449; Web site: http://www.aace.org.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Computer Uses in Education; *Educational Resources; Educational Technology; Foreign Countries; Higher Education; Information Networks; Information Scientists; Instructional Materials; *Online Catalogs; Open Universities; *Shared Resources and Services
IDENTIFIERS	Europe

ABSTRACT

The European project, UNIVERSAL--Universal Exchange for Pan-European Higher Education, is an attempt to demonstrate an open exchange of learning resources (LRs) between higher education institutions across Europe. The goal is to create and manage an open market of learning resources by introducing a brokerage platform with a standard way of describing the pedagogical, administrative, and technical characteristics of learning resources. The system will enable institutions to enrich their curricula with remotely sourced material. It is intended to be compatible with a variety of similar models pursued by different institutions, including open universities and alliances between peer institutions. The business-to-business oriented brokerage will embrace offers, enquiries, booking, and delivery of LRs. This paper illustrates how the issue of providing a user-friendly access mechanism to learning resources has been addressed by UNIVERSAL. The focus is devoted to catalogue design: its creation, development, structure, and organization. Attention is also given to the automated categorization of learning resources within the category tree, creation, and maintenance of the catalog hierarchy. The described approach enables gradual automation of the categorization process: from manual to semi-automated, and finally to a fully automated categorization process. (Contains 14 references.) (AEF)

Reproductions supplied by EDRS are the best that can be made
from the original document.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

G.H. Marks

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Learning Resource Catalogue Design of the UNIVERSAL Brokerage Platform

Gorazd Vrabic
Jozef Stefan Institute
Laboratory for Open Systems and Networks
Ljubljana, Slovenia
gorazd@e5.ijs.si

Bernd Simon
Vienna University of Economics and Business Administration
Department of Information Systems, New Media
Augasse 2-6, A-1090 Vienna, Austria
bernd.simon@wu-wien.ac.at

ED 466 227

Abstract: This paper illustrates how the issue of providing user-friendly access-mechanism to learning resources has been addressed by UNIVERSAL, an open, pan-European brokerage platform for learning resources. The focus is devoted to catalogue design; its creation, development, structure and organization. More attention is also given to the automated categorization of learning resources within the category tree, creation and maintenance of the catalogue hierarchy. The described approach enables gradual automation of the categorization process: from manual to semi-automated and finally to a fully automated categorization process.

Introduction

The European project "UNIVERSAL - Universal Exchange for Pan-European Higher Education" is an attempt to demonstrate an open exchange of learning resources (LRs) between higher education institutions across Europe. The goal is to create and manage an open market of learning resources by introducing a brokerage platform with a standard way of describing the pedagogical, administrative and technical characteristics of learning resources. The system will enable institutions to enrich their curricula with remotely sourced material. It is intended to be compatible with a variety of similar models pursued by different institutions, including open universities and alliances between peer institutions. The business-to-business oriented brokerage platform will embrace offers, enquiries, booking and delivery of LRs.

Based on the IEEE Learning Objects Metadata (LOM) standard, four aggregation levels of LRs are introduced: course, unit, lesson, and fragment. A course is defined as a set of units or lessons contributing to one learning goal and lasting not longer than one semester or term. Courses can be organized in units, which in turn consist of individual lessons. A full course description does not always have to consist of units, it can also be based on lessons. A lesson is part of a course or a unit, which does not necessarily have to be offered in UNIVERSAL as well. Lessons can consist of multiple fragments, which are also referred to as supporting material (Guth et al. 2001).

The platform supports synchronous as well as asynchronous material. The following examples of LRs are taken from the preliminary version of the UNIVERSAL catalogue:

- A course dealing with Icelandic Sagas (aggregation level: course; delivery mode: synchronous);
- A web-based training application instructing ophthalmologists on how to diagnose patients (aggregation level: unit; delivery mode: asynchronous);
- A recorded session of a course in international marketing dealing with a Levi Strauss case study (aggregation level: lesson; delivery mode: asynchronous);
- PowerPoint slides, exercises, and a term project description supporting an introductory course in information technology (aggregation level: fragment; delivery mode: asynchronous).

LR metadata is maintained by the UNIVERSAL metadata engine combining the eXtended Markup Language (XML) and the Resource Description Frameworks (RDF) approach. The RDF Schemas used for describing LRs and LR related information such as authors, delivery systems, taxonomies, etc. are available at http://nm.wu-wien.ac.at/universal/metadata. The XML:RDF approach used ensures scalability and transferability of UNIVERSAL

IR021298

BEST COPY AVAILABLE

repository of LR metadata instances (Guth et al 2001), because mapping the Universal metadata schema with LOM or Dublin Core can easily be carried out by generating LOM compliant or Dublin Core compliant metadata instances.

A key innovation of the UNIVERSAL project is the tight integration of the LR catalogue and the delivery systems. A delivery management engine prepares the grounds for the delivery of LRs by granting access rights and verifies the availability of the content. A thin generic interface layer provides communication functionality between brokerage and delivery system, such as availability checks, authentication and authorization services, delivery negotiation and delivery supervision. UNVERSAL will offer a fully implemented delivery interface to the following restricted set of delivery systems, but is also open to others by providing a generic delivery interface:

- Apache web server (asynchronous packaged content);
- Hyperwave's E-learning Suite (learning management system);
- RealNetwork's Realserver (asynchronous streaming content);
- Isabel (synchronous collaboration tool and video conferencing system).

Requirements for Catalogue Design

The UNIVERSAL catalogue is an essential part of the system, since it brings demand and supply together and has direct implications on the costs of deploying the system. Considering the effected user groups (potential LR consumers, LR providers, and platform administrators), several issues arise. On the one hand catalogue design has to assure that potential LR consumers have efficient access to the LR they are interested in. It is essential for the potential consumer to find a desired LR quickly and with ease, since high search costs reduce the perceived net-value of a web-based information system (Kaukal & Simon 1999). On the other hand, platform hosts are interested in keeping the catalogue maintenance costs at a minimum level. Brin and Page argue that human maintained catalogues are expensive to build and maintain (Brin & Page 1998). In this paper we present a scalable approach that should make up this drawback.

The catalogue requirements from a LR provider's perspective are based on a qualitative survey at Austrian universities. Seven higher education experts, managers or pioneers in applying new media in the field of traditional higher education have been asked about potential obstacles for using systems such as UNIVERSAL. Concerning catalogue design the survey revealed that faculty members are reluctant to provide LRs of high quality when the quality of the listed LRs is not of a similar level. There is also a need of accessing LR-specific assessment and usage data in order to charge LRs depending on their quality.

A catalogue must have structured content of a well thought and well-defined standard set of fields that can be extended to all metadata elements to accommodate the uniqueness of the different types of resources as the number of catalogued resources increases and as the category tree expands.

The content of a catalogue must answer the following questions:

- What is available on the subject of user's area of interest (AOI)?
- What is the quality of available LRs and what do they look like?
- What is the procedure for the user to get a specific LR?

The Categorization Process

A hierarchical structure of a catalogue in the UNIVERSAL project will be presented in a multi-level hierarchy. The first level is a subject-centric level with main thematic categories which provide the subject structure of the catalogue content based on keywords from a selected classification schema, such as the Dutch Basic Classification (available at http://www.konbib.nl:88/kb/resources/frameset_kb.html?/kb/vak/basis/bc98-en.html).The top division of the classification schema is of particular importance, since it provides the first entry point to the catalogue and effects the organization of all the following categories. The second level is a title/authors/publication time/etc.-centric level where the structure is based on a certain elements of the LR meta-data instances. These attributes are used for building categories (title/authors...- the middle division). These elements (categories) are selected by the catalogue administrator. As the number of LRs grows the number of categories used in the middle division will expand. The second level may be repeated several times with a different category in each level (repeated middle division). The next level is a type-centric level with four types of LRs: course, unit, lesson and fragment (actually it is a specific type of middle division). The last (the bottom division) is for the actual links to

resources with all the important information for every resource, respectively. The top and middle division(s) are essential for navigation around the catalogue, and the bottom division holds information where the content of the catalogue is found.

The goal of the categorization process is to find the most appropriate categories under which every LR should be classified. The categorization process has two inputs. On the one hand, XML:RDF metadata instances of LRs are exploited and, on the other, the pre-existing category tree in a catalogue (let us assume that an initial version of catalogue exists) within which every LR must be categorized is taken under consideration (Fig. 1).

The category tree consists of a tree with a description for every node, which identifies the category and helps to properly categorize every new LR, which is to be added to the catalogue. The output of the categorization is a vector of weights associated to each node in the category tree. Each element of that vector represents the probability that the resource should belong to the category represented by each node. The vector of weights will be determined by using natural language processing tools, which will take advantage of a UNIVERSAL ontology

The classification algorithm considers each node in the category tree as a context path. The weights from all context paths of the same resource are added (sum of weight vectors is performed) and if the normalized sum for a certain node is greater than a certain threshold, the resource is classified under that node. The mechanism allows classifying a resource under more than one node. However, the same resource could never be classified in nodes, which are descendant of one another. The title of a category is not sufficient to fully identify the category. The whole path (context path) is necessary to disambiguate among the categories.

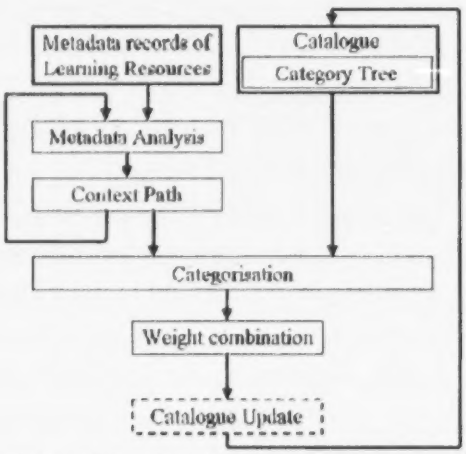

Figure 1: Building a catalogue hierarchy based on categorization. Categorization exploits metadata instances of LRs and category tree in a catalogue (double rectangles). The final outcome of the whole process is "catalogue update" (dashed rectangle).

The categorization algorithm works as follows: it first computes candidate context paths for classifying a resource, then it selects among candidates, and finally it updates the category tree and the entire catalogue. Every context path in a category tree (path from the root node to each node on every level in a category tree) is represented by a vector of weights $C_1, C_2,\ldots,C_n$, where n is the number of levels in a context path. Each weight presents the probability that the resource should belong to the node. The initial value of every weight is 0. The values from the metadata instance are searched for an identifying description. If the value is found in node j (j = 1, 2...n), a certain discrimination weight is added to the weight Cj. Any path containing non zero elements is considered as a potential candidate. The selection among these candidates is performed as follows. Any path with zeros in its leading fields should be discarded. This means that matches are found only in some subcategory, but not in the top-level categories. If more candidates remain then a candidate with longer path should be selected. This forces categorization under the most specific category. The selected candidate paths associated to the resource are stored in the database. If the same LR could be classified into different categories with different context paths (the same resource is reached from a different path), the categorization mechanism should either enforce the indication of the category for a resource or should suggest alternative categories for the resource (creation of a new category).

There are two possible cases when the categorization process should extend the categories in the catalogue. The first one is when the number of entries of a category exceeds a certain threshold (e.g. 40 entries). In this case the

list of LRs would become too long for displaying it in a user-friendly manner. Re-classifying those LRs in new subcategories would significantly support finding desired LR effectively. The second case is when an alarming number of resources do not find a proper place in the catalogue hierarchy. Then a new category should be created.

To extend the categories, the context paths should be analysed further (Maarek & Shaul 1996). In both cases, the context path in a single large category should be searched for elements in the metadata, which produce partitions of the resources. All unused elements from the metadata instances of resources in that part of category tree should be considered as candidates for creating new categories.

Since several partitioning alternatives may arise, statistical or neural analysis techniques should be applied to rank the most promising alternatives and present them to the catalogue administrator who will decide which subcategories to add to the catalogue. The experienced catalogue administrator could do the selection of an adequate category instead of applying automated statistical or neural analysis techniques. However, we can already find products such as Autonomy and Vistabar on the market that are designed to automate the process of categorization. The choice between the manual or automated categorization lies in the decision of the level of automation of the whole cataloguing process (Salton et al. 1994).

User Interface Design of the UNIVERSAL Catalogue

Fig. 2 illustrates the two different types of user access to a repository of LRs: browsing and searching (Manber et al. 1996). A user starts browsing through catalogue and wants to discover what is available on his AOI. The top-level category (root node) is the starting point in his attempt to find a resource meeting his general idea of the item of interest. Then in the information discovery process user continues to browse and progresses to more and more details by refining the selection in every node moving from higher to lower level to the last possible selection. Browsing mechanisms enable users to locate and identify their AOI. At the same time a user can view relationships such as overlapping, intersecting and containment of the available resources with their AOI. A second way to explore the content of a catalogue is by defining an AOI with the search query. This process could be repeated several times and every time with a different function: search or browse. Therefore a list of resources from the search function must also contain information in which node of the category tree a resource can be found.

Catalogue Browsing Interface

The UNIVERSAL catalogue will provide a structured view on the LRs offered through the UNIVERSAL brokerage platform by introducing multiple levels of categories. For every node in a category tree index will be unique. An index is understood as a list of LRs sorted in a specific way (most common is descending alphabetical order) (Srihari 1995). Each index page should present its hierarchical location in the catalogue. The user interface should allow users to click on any node in the hierarchy (when a user is visiting certain node, the whole context path should be visible as array of hyperlinks) and when clicked it should send a user directly to that catalogue node. Which nodes (hyperlinks to nodes) should be visible (click-able) from certain node is a matter of discussion. However, the rule must be the same for all nodes (for example: all nodes in a context path and first descendant level to that node - if existing).

The catalogue in UNIVERSAL will use a hierarchical arrangement. In other words, users will be able to "drill down" from general to a more detailed level within the catalogue. Here a hierarchy may be presented as a triangle, with the root node at the top corner, and the lowest level of detail as the bottom edge of the triangle. If the hierarchical structure of this presentation is close to an equal-sided triangle then the hierarchy is assumed to be sensible and adequate. Smaller catalogues usually need only a few levels, while bigger may have a middle level of nodes between the root node and the lower level. As one same resource can be assigned to more than one catalogue node the same LR can be found also in several catalogue nodes of the category tree (if selected elements from the metadata instances allow categorization of the same resource into those nodes).

Search Interface

When search is performed the user searches a catalogue with a specific, focused goal. The search interface of the UNIVERSAL brokerage platform will provide the input dialog or selection fields to enter the specifics. The search could be performed "globally", i.e. within the complete repository of LRs. However also "local" search will be supported. Here only the content of that part of the catalogue, where the user is browsing will be subject of the search (this includes all nodes bellow a certain node). The search results are returned as a list of LRs. Then, the user

can explore the details of the returned matches through further selections, query refinements, and viewing of all information relevant to meet the initial specifications as closely as possible (Chalmers 1998). It is planned that the LR search engine will support previews and users will be able even to download a sample set of the resources to determine their match to the AOI.

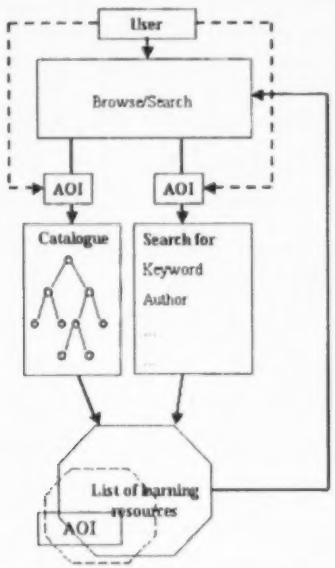

Figure 2: A user may access resources in two ways: by browsing through the catalogue or by searching mechanism. Users are looking for offered resources meeting their Area of Interest (AOI). A link between the area of interest and the catalogue shows resources available in the catalogue meeting their AOI. The result of a Browse/Search activity is a list of resources (Browse/Search activity can be performed further on this list of resources). List of resources in every step is different moving from one position/size to another (dotted stop sign shape square) with goal to match with AOI square as much as possible.

The search engine of the UNIVERSAL brokerage platform will perform search based on the descriptions of resources (content of metadata instances) and will provide results as a linear list of resources. This is unsatisfactory because the list can be quite long, and indication of possible grouping of related material is essential. Splitting the list of resources into categories and displaying hits based on catalogue categories will significantly facilitate selecting those LRs the user is most likely interested in.

The scope of a search in a UNIVERSAL catalogue will be limited by the user's search location within the category tree. To search over one category a user should go to this category node. To search over whole catalogue (all LRs), he should simply go to the top-level node (root) and search from there. Therefore the user in order to broaden or narrow his search scope has to go to a more appropriate node in the catalogue.

When search will be performed a list of LRs will be displayed. A user is usually interested in how this list is near his AOI. The title of the LR is often not enough for a user to identify if a LR is on the subject of user's interest. The following information will be available: language, title, aggregation level, learning goals, educational aims, description, author, keywords, duration, costs, copyright restriction and categorization within the category tree. Depending on the LR aggregation, additional information is provided:

- For fragments: learning resources type (lecture, case study, exercise, simulation...), format, version, status (draft, final, revised);
- For lessons: instructional design (directed learning, self directed learning, collaborative work), starting time, ending time and time zone (for live lessons), delivery platform specifications;
- For units: pre-requisites (knowledge, experience), instructional design (see above);
- For courses: pre-requisites, program (undergraduate, graduate, postgraduate).

Conclusions

There are many approaches for building a catalogue hierarchy described in literature (Hahsler & Simon 2000). The proposed approach in this document is taking advantage of highly structured and highly flexible repository of metadata instances achieved by combining the advantages of XML and RDF (Guth et al. 2001) and enables full automation of a cataloguing process. This has an impact on fast and objective categorization of LRs into categories avoiding subjective estimation of the catalogue administrator or users in which category specific LR should be found (Cleverdon 1984). This approach also manages fully automated growth of catalogue and category tree as the number of LRs offered through UBP increases. Moreover, it facilitates a gradual automation of the whole categorization process. In an early stage of the project manual categorization will be done. Planned stages are:

- Fully manual categorization and manual managing of a catalogue (expansion of a category tree);
- Automated categorization in an already existing category tree and manual expansion of a category tree;
- Automated categorization of LRs and automated expansion of a category tree by applying analysis techniques of metadata for further partitioning of categories.

References

Autonomy, http://www.topic.com.au/products/autonomy.html.

Brin, S., Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems, 30, 107-117.

Chalmers, M., Rodden, K., Brodbeck, D. (1998). The order of things: activity-centred information access. Computer Networks and ISDN Systems, 30, 359-367.

Cleverdon, C. (1984). Optimizing convenient online access to bibliographic databases. Information Services and Use, 4, 37-47.

Guth, S., Neumann, G., Simon B. (2001). UNIVERSAL - Design Spaces of Learning Media. Proceedings of the 34th Hawaii International Conference on System Sciences 2001, Maui, USA.

Hahsler, M., Simon, B. (2000). User-centered Navigation Re-Design for Web-based Information Systems. Proceedings of the 6th Americas Conference on Information Systems 2000, Long Beach, USA, 192-198.

IEEE Learning Technology Standards Committee (LTSC), http://ltsc.ieee.org/.

Kaukal, M., Simon, B. (1999). Redesign von WWW-basierten Masseninformationssystemen, Electronic Business Engineering, Heidelberg, Germany, 51-66.

Maarek, Y. S., Shaul, I. Z. B. (1996). Automatically organizing bookmarks per content. Computer Networks and ISDN Systems, 28, 1321-1333.

Manber, U., Gopal, B., Smith, M. (1996). Combining browsing and searching. W3 Distributed Indexing/Searching Workshop 1996, MIT, Cambridge, Massachusetts, USA.

Salton, G., Allan, J., Buckley, C., Singhal, A. (1994). Automatic analysis, theme generation, and summarization of machine-readable text. Science, 264, 1421-1426.

Srihari, R. K. (1995). Automatic Indexing and Content-Based Retrieval of Captioned Images. Computer, 28 (9), 49-56.

Technical Annex of the UNIVERSAL project, IST-1999-11747.

Vistabar, http://research.compaq.com/SRC/WebArcheology/Vistabar.html.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

☑ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").